DOCUMENT RESUME

ED 190 075 IR 008 556

AUTHOR Harrison, Isom, Jr.
TITLE CAS77 and CAS7276: A Review.
PUB DATE 28 Nov 79
NOTE 15p.: Paper presented at the National Online Information Meeting (New York, NY, March 25-27, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Chemistry; Data Bases; *Information Retrieval; Information Systems; *Online Systems; *Search Strategies

ABSTRACT

This paper describes the content, organization, specifications, and methods of use of the CAS77 and CAS/276 online files of worldwide chemical literature, databases produced by Chemical Abstracts Service and available from System Development Corporation (SDC). The scope of the databases, their unit record, their data elements, their modes of subject searching, available user aids, and charges to users are detailed. Tables identify directly searchable data elements provided by SDC and database specifications for the CAS77 file. Figures illustrate samples of a unit record, a CHEMDEX file search, and a citation resulting from a search. (FM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

PREPRINT UCRL- 82718

Lawrence Livermore Laboratory

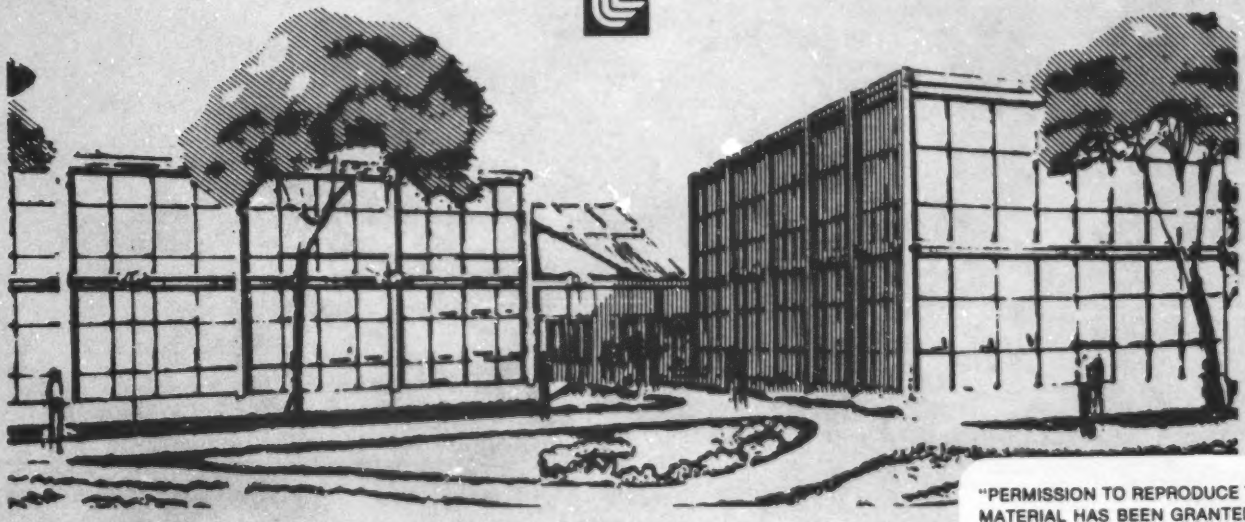
CAS77 AND CAS7276: A REVIEW

Isom Harrison, Jr.

November 28, 1979

This paper was prepared for the First National Online Information Meeting, Sheraton Hotel, New York City, March, 1980.

This is a preprint of a paper intended for publication in a journal or proceedings. Since changes may be made before publication, this preprint is made available with the understanding that it will not be cited or reproduced without the permission of the author.



"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Nan C. Prentice
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED190075

CAS77 AND CAS7276: A REVIEW*

Isom Harrison, Jr.

Lawrence Livermore Laboratory, University of California

Livermore, California 94550

ABSTRACT

The content, organization, specifications, and methods of use of the CAS77 and CAS7276 online files of worldwide chemical literature are described. Specific examples of a record and a search are given. We have found at Lawrence Livermore Laboratory that even sophisticated literature searches can be made by non-technical personnel as a result of certain attributes of these systems: the many access points to the data that are provided by System Development Corporation and the excellent abstracting and indexing service provided by Chemical Abstracts Service.

A comparison of a search done in ORBIT III and ORBIT IV is also given along with a list of the searchable categories available in the chemical files.

*This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore Laboratory under contract No. W-7405-Eng-48.

INTRODUCTION

The purpose of this paper is to point out some of the uses, by both technical and non-technical personnel, and some advantages of the CAS77 and CAS7276 online files of worldwide chemical literature.

DESCRIPTION AND USE OF FILES

Scope

The CAS77 and CAS7276 files are available online from System Development Corporation (SDC).* The databases are produced by Chemical Abstracts Service (CAS), American Chemical Society. The files cover worldwide chemical science literature, including journals, conference proceedings, new books, patents from 26 countries, and government research reports. The files have been divided by SDS into 80 sections called file segments. Biochemistry is included in Sec1-Sec20; organic chemistry, Sec21-Sec34; macromolecular chemistry, Sec35-Sec46; chemical engineering, Sec47-Sec64; and physical chemistry, Sec65-Sec80.

The CAS77 file covers 1977 to date and is updated biweekly. The database is a combination of Chemical Abstract (CA) condensates and CASIA (Chemical Abstracts Subject Index Alert) and covers the tenth collective index period. The CAS7276 file covers 1972-1976 and covers the ninth collective index period.

The data base sizes are approximately 400,000 records per year of coverage for CAS77 and approximately 1,800,000 total for CAS7276.

---

*Reference to a company or product name does not imply approval or recommendation of the product by the University of California or the U.S. Department of Energy to the exclusion of others that may be suitable.

The Unit Record

The unit record consists of assigned vocabulary terms (from CASIA) and citations (see Fig. 1). The citation includes an accession number, title, author, corporate source, source, document type (e.g., P = patent, J = journal, B = book), language, category codes (CA section numbers), index terms (which include registry numbers), and supplementary terms. The citation does not include abstracts. The unit record categories may vary with the type of original documents (e.g., book vs patent, or book vs journal). For a patent document, the patent number, classification, country, and priority country are included in the source category.

Data Elements

Many data elements are available to users of the CAS77 and CAS7276 data bases. Searches can be very broad or very specific depending on which data elements the searcher chooses to use. Data elements can be used in combination when searching for a particular topic. For example, category codes can be used in conjunction with index terms to ensure a more thorough search on a topic. Another example is the use of registry numbers for a compound in conjunction with the compound name or with synonyms for the compound.

The many, directly searchable, data elements provided by SDC allow for many access or entry points to the data available in the two files. These numerous access points allow even the novice user with little or no technical background to do simple searches. As a memory jogger, the directly searchable categories are listed in Table 1.

At the Lawrence Livermore Laboratory (LLL), we have found that technical information specialists with a non-technical background can do very sophisticated searches in the CAS files because of the excellent abstracting

and indexing done by CAS and the many access points provided by SDC to the data elements. The searchers, of course, know the ORBIT system and use the ORBIT Quick Reference Guide frequently.

Subject Searching

The Research Information Group of LLL uses the CAS77 and CAS7276 files for several reasons. One is for "quick and dirty" searches asked for at the reference desk. These searches may come in the form of an author search, substance search, reference verification, or subject search. They usually take only a few minutes to complete, and the customer can get references immediately. We also use the files for current awareness, retrospective bibliographic searching, and for bibliography preparation.

In the past (in ORBIT III), subject terms in the title could not be searched directly. The term would have to be searched in the index term category, and then the resulting hits had to be stringsearched for the title terms (e.g., STRS 1 (TI) :Blood Analysis:). If the stringsearch was for title terms only, the qualifier TI was not necessary, because the default category for stringsearch was the title category. However, with the birth of ORBIT IV, title terms are now directly searchable. Term A/Ti and Term B/Ti specify that Term A and Term B, respectively, must be in the title.

Another alternative to title searching is a search of the basic index category. This category contains terms taken from the titles, index terms, original titles, and supplementary terms. After searching for subject terms in the basic index, the searcher can then do a stringsearch for the terms in the title. However, because this procedure is more time consuming, it is less cost effective. Some searchers may choose to use the basic index category

because of the specificity that a title search requires. Some good hits may be lost by restricting the search to the title terms only.

At LLL we frequently need to search for complex chemical names. In ORBIT III, complex chemical names found in titles were indexed by SDC so that they could be retrieved with initial numerics with commas followed by a hyphen and an alphabetic string. For example, a compound like 9,9-dibromobicyclo(6.1.0) nonane could be entered several ways: 9,9-dibromobicyclo, 99dibromobicyclo, nonane, or dibromobicyclo in combination with nonane. In ORBIT IV files (CAS77 and CAS7276), if we enter 9,9-dibromobicyclo (or 99bromobicyclo), we get a no postings message, because the term is not indexed as a subject heading. To retrieve 9,9-dibromobicyclo(6.1.0)nonane in ORBIT IV, we would have to use the registry number for that substance. To obtain the registry number for the substance, we must consult the online index of chemical substances, called CHEMDEX (CHEMNAME on DIALOG). CHEMDEX contains registry numbers, molecular formulas, CA name synonyms, and ring system information. Once the registry number has been obtained from CHEMDEX, the searcher can log into the CAS77 (or CAS7276) file, enter the registry number (see Fig. 2), and obtain a unique number of postings for that substance (see Fig. 3).

The addition of CASIA to the chemistry files on SDC greatly enhanced the searching capabilities. The use of controlled vocabulary terms and registry numbers virtually assures that the searcher will get good hits with few or no false drops.

A summary of the data base specifications for CAS77 is given in Table 2.

User Aids Available

Several user aids are available from both SDC and CAS for searching CAS77 and CAS7276. In addition to the User's Manual and the Quick Reference Guide,[1]

SDC has developed a microfiche with alphabetical subject-term listing, showing the number of postings available for each term in the data base. It sells for $60 (may change with time) and is available from the SDC Search Service (see footnote to Table 2).

The following user aids, which are invaluable for searching the combined CA condensates CASIA files, are available from CAS: the printed CA Index Guide,[2] CA General Subject Index Headings List,[3] CA Taxonomic Subject Index Headings List,[4] Chemical Substance Name-Registry Number List,[5] CA Nomenclature Word Frequency List,[6] and the Key-Letter-In-Context (KLIC) Index of CA Nomenclature Terms.[7] The last five items are available on microform. Also available from CAS is the booklet "Subject Coverage and Arrangement of Abstracts by Section in the Chemical Abstracts."[8]

Price

The CAS77 and CAS7276 files are available online from SDC. The charge for computer connect time is $60/hour and $.12 for each citation printed offline. Questions regarding online searching of CAS77 and CAS7276 should be directed to SDC's Action Desk.

CONCLUSIONS

The CAS77 and CAS7276 files on chemical literature provide distinct advantages for information searches, even by non-technical personnel, because of their large data base, the many access points provided to the available data, and the excellent abstracting and indexing service furnished.

REFERENCES

1. ORBIT User's Manual and ORBIT Quick Reference Guide (System Development Corporation, Santa Monica, CA, 1979).
2. CA Index Guide, 9th Collective Index (Chemical Abstracts Service, Columbus, Ohio, 1972-76).
3. CA General Subject Index (Chemical Abstracts Service, Columbus, Ohio, 1977-78).
4. CA Taxonomic Subject Index Heading List (Chemical Abstracts Service, Columbus, Ohio).
5. CA Chemical Substance Index (Chemical Abstracts Service, Columbus, Ohio, 1977-78).
6. CA Nomenclature Word Frequency List (Chemical Abstracts Service, Columbus, Ohio).
7. Key-Letter-In-Context Index of CA Nomenclature Terms (Chemical Abstracts Service, Columbus, Ohio).
8. Subject Coverage and Arrangement of Abstracts by Section in Chemical Abstracts, 2nd edition (Chemical Abstracts Service, Columbus, Ohio, 1975).

NOTICE

"This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Department of Energy, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately-owned rights."

YDB:jm

TABLE 1. Directly searchable data elements provided by SDC.

Category (qualifier[a])	Category (qualifier)
Basic index (no qualifier needed)	Location (LO)
Index terms (IT)	Patent/priority number (PN)
Authors (AU)	Patent classification (PCL)
Organizational source (OS)	Patent country (PC)
Category codes (CC)	Priority country (PR)
File segment (FS)	International Standard Serial No. (ISSN)
Language (LA)	Document type (DT)
City (CI)	Journal codes (JC)

[a]In subject searching, a slash precedes each qualifier (e.g., BUTADIENE/TI).

TABLE 2. Data base specifications for CAS77 file, which uses Chemical Abstracts as corresponding print publication.[a]

Item	Specifications
Subject matter in file	Biochemistry, organic chemistry, macromolecular chemistry, applied chemistry, chemical engineering, physical and analytical chemistry
Source documents	Journals, patents, books, conference proceedings, government research reports
Size of data base	1,800,000 records for CAS7276; 400,000 records per year of coverage for CAS77
Language of source document	All languages
Language of data base	English
Backfiles online	1972 to present
Systems on which files are available	SDC ORBIT IV
Code used to call up data base	TYMNET, TELENET
Transmission speeds	10, 15, 30 and 120 CPS
Service available	Worldwide
Price per connect hour	$60
Prior conditions for user fees	None
Restrictions on who may use it	None
Tapes available	Yes
Frequency of updates	Biweekly
Type of record	Citation
Database searchable in free text	Yes
Thesaurus available	Index Guides

TABLE 2. (Continued)

Item	Specifications
Other search aids available	User Manual, Quick Reference Guide
Offline prints available	Yes
Microform of original documents available	Yes

[a]Contacts for further information:

Chemical Abstracts Service (American Chemical Society), P.O. Box 3012, Columbus, Ohio 43210 (Phone (614) 421-6940).

System Development Corporation Search Service, 2500 Colorado Avenue, Santa Monica, CA 90406 (Phone (800) 352-6689 for California or (800) 421-7229).

PROG:

-1-

AN - CA86 - 80(1)

TI - Effects of mercury compounds in tuna on organisms. II. Accumulation of mercury and selenium in cats after long-term administration of tuna.

AU - Shiramizii, Michiko, Yamaguchi, Seiya, Kaku, Shunsuki

OS - Kurume Univ., Sch. Med., Kurume, Japan

SO - Sangyo Igaku (SAIGBL), V. 18 (2), p. 123-35, 1976

DT - J (journal)

LA - Japan

CC - SEC4-3; SEC17

IT - 7782-49-2, biol study; 7439-97-6, biol study: (or organs of cats, after tuna fish ingestion).

IT - Hair; blood: (mercury of, after tuna fish ingestion).

IT - Brain, metabolism, Kidney, metabolism; Liver, metabolism: (mercury and selenium distribution in, after tuna fish ingestion).

ST - Organ

FIG. 1. Sample of a Unit Record.

```
     SS 1/c?
USER:
     9,9-Dibromobicyclo:
PROG:
     MM (9,9-Dibromobicyclo:) (3)
        1.  9,9-Dibromobicyclo(6.1.0)Non-2-En/N
        2.  9,9-Dibromobicyclo(6.1.0)Non-4-En/N
        3.  9,9-Dibromobicyclo(6.1.0)Nonane/N
SPECIFY NUMBERS, ALL OR NONE
USER:
     3
PROG:
     SS 1 PSTG (1)
     SS 2/c?
USER:
PRT FULL
PROG:
-1-
     RN - 1196-95-8
     MF - C9H14Br2
      N - Bicyclo(6.1.0)Nonane, 9,9-Dibromo
    RSD - 2 Ring(s), C3, C8
   SYNN - 9,9-Dibromobicyclo(6.1.0)Nonane
SS 2/C?
FILE CAS77
PROG:
ELAPSED TIME ON CHEMDEX:  0.03 Hrs.
YOU ARE NOW CONNECTED TO HE CAS77 DATABASE COVERS 10TH COLLECTIVE PERIOD
FORWARD:
SS 1/c?
USER:
1196-95-8/RN
PROG:
SS 1 PSTG (5)
SS 2/c?
USER:
PRT FULL 1
```

FIG. 2. Sample of CHEMDEX file search.

PROG:

-1-

AN - CA90-103461(13)

TI - Chemistry of Gem-Dihalocyclopropanes. XV. Reactions of Gem-Dihalocyclopropanes with sodium Bis(2-methoxyethoxy) Aluminum hydride, lithium aluminum hydride, and sodium borohydrice.

AU - Sydnes, Leiu K.; Skatteboel, Lars

OS - Univ. Oslo, Dep. Chem., Oslo, Norway

SO - Acta Chem. Scand., Ser. B (ACBOCV), V B32 (9) p. 632-8, 1978, ISSN 03024369

DT - J (JOURNAL)

LA - ENG

CC - SEC24-2; SEC22

IT - 1196-95-8; 2415-79-4; 2415-80-7; 3135-31-7; 3234-51-3; 6998-44-8; 58683-49-1; 65655-79-0; (REDN. OF, MECHANISM OF)

IT - 2568-73-2; 52100-83-1; 52100-86-4; 58987-26-1; 65655-78-9: (REDN. OF)

IT - 16853-85-3; 16940-66-2; 22722-98-1: (REDN. BY, OF GEM-DIHALOCYCLOPROPANES, MECHANISM OF)

IT - 64670-28-6; 67074-45-7: (PREPN. AND REDN. OF)

IT - REDUCTION: (OF GEM-DIHALOCYCLOPROPANE: MECHANISM OF)

ST - CYCLOPROPANE: DIHALO: HALOCYCLOPROPANE: BICYCLONONANE

FIG. 3. Citation resulting from search shown in Fig. 2.